DOCUMENT RESUME

ED 121 742 SP 010 003

AUTHOR Gay, L. R.
TITLE Evolution of an Innovation: Development of a Competency-Based Educational Research Course, 1972-1976.
PUB DATE Apr 76
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS Credit No Credit Grading; *Educational Research; *Education Courses; Feedback; Higher Education; Individualized Programs; *Performance Based Teacher Education; Performance Criteria; *Teacher Education

ABSTRACT

During the fall quarter 1972, a total competency-based educational research course was offered at Florida International University. At that time, the course was characterized by prescribed competencies, criterion-referenced evaluation, the absence of tests, credit-no credit grading, self pacing, unlimited recycling to criterion, and individual feedback. During the past three and one-half years the course content has remained relatively stable, but course procedures have changed. Most of the changes have been in the area of student evaluation and have been based on student performance data, student and instructor feedback, and the results of experimental research conducted within the course. It has now been determined that (1) prespecified competencies and criterion-referenced evaluation are essential and desirable components of a competency-based research course; (2) testing of subcompetencies is apparently essential in order to assure comprehension of prerequisite knowledge; (3) credit-no credit grading is unacceptable on both philosophical and practical grounds to students, instructors, employers, and graduate schools; (4) self pacing is detrimental to both students and instructors; and (5) unlimited recycling is inefficient and unproductive, even when accompanied by intensive, individualized feedback. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121742

EVOLUTION OF AN INNOVATION:

Development of a Competency-Based Educational Research Course, 1972-1976

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

L. R. Gay

Florida International University

Miami, Florida

Paper Presented at the American Educational Research Association Convention

April, 1976

SP 010 003

CONTENTS

	Page
Introduction	1
Prespecified Competencies	2
Criterion-Referenced Evaluation	2
Testing	3
Grading	6
Pacing	9
Recycling	10
Conclusions	13
APPENDIX A	15
APPENDIX B	18

LIST OF FIGURES

Page

Figure 1 Percentage of Students completing a competency-based research course by the end of each of 11 quarters (excluding Summers), Fall, 1972-Winter, 1976.....11

During the Fall Quarter of 1972, a totally competency-based educational research course was offered for the first time at Florida International University (FIU). Considerable effort was involved in course development including the preparation of a modular course guide and the assemblage of research reports and models of expected student performance. At that time, the course was characterized by the following: prespecified competencies, criterion-referenced evaluation, the absence of tests, credit-no credit grading, self pacing, and unlimited recycling to criterion, accompanied by considerable individual feedback.

During the past three and one-half years, at least one section of the course has been offered every quarter. While course content has remained relatively stable, course procedures have changed in a number of ways. Most of the changes have been in the area of student evaluation and have been made based on student performance data, student and instructor feedback, and the results of experimental research conducted within the course. Although a number of the changes have contributed to increasing the manageability of the course, not one change was made solely in the interest of manageability. In other words, no revision was instituted to the detriment of students for the purpose of making procedures more manageable for instructors;

all changes have either positively affected student performance or have had no effect.

Prespecified Competencies

The course competencies (tasks) and subcompetencies (enablers) have remained relatively stable over the years. Basically, the only changes have been the addition or deletion of enablers related to certain tasks. The stability of the tasks and enablers is probably due mainly to the manner in which they were originally formulated. A list of essential skills for consumers and producers of educational research (with emphasis on the former) was compiled. A Gagné-type task analysis was used in identifying knowledge and skills prerequisite for the accomplishment of each task. Since the research course is required of all master's degree students in education at FIU, the resulting tasks and enablers were then reviewed and critiqued by members of all units in the School of Education. The finalized set of competencies and subcompetencies then formed the basis of the modular course guide.

After nearly four years, students still express gratitude for their right to know, from the first day of the course, exactly where they are going and what is expected of them.

Criterion-Referenced Evaluation

Evaluation has always been based on the degree to which students meet criteria rather than on a norm-referenced

comparison of student performance. Improvements have been made, however, with respect to the ways in which expected performances are communicated to students. It has been observed that concrete models, rather than verbal or written explanations, result in superior performance. The most effective models are those produced by other students as opposed to examples taken from published research. As an example, a student can better understand and, more importantly, identify with a review of related literature written by a former student. Psychologically, student-produced models cause students to believe that expected performances are within their capabilities. Further, since journal-quality products are not expected of students in an introductory research course, examples of other students' work actually provide truer models of expected level of performance.

Testing

Originally, the course involved no tests. All enablers were submitted as written exercises.[1] At the end of the 1972-73 school year, the sentiment among the course instructors was that this approach to enabler evaluation was unsatisfactory for several reasons. The major one was that many students could apparently complete a set of written enablers in a semi-automatic manner, with very little comprehension of

---

[1]Since tasks involve demonstration of the ability to perform research functions, such as the ability to select and apply an experimental design appropriate for a given hypothesis, "testing" situations do not represent a viable alternative approach to evaluation of tasks.

what they had written. Secondly, since many enablers do not require individualized responses (as do tasks), these were apparently a number of opportunistic students who were simply copying enabler responses from their more ambitious friends; this tactic also led to an obvious lack of comprehension. Consequently, wherever feasible, enabler tests were developed to measure required knowledge and skills.[2]

The introduction of tests during the Fall Quarter of the 1973-74 school year created a heated controversy involving both instructors and students. For nearly a year the pros and cons of testing were debated. Some instructors felt that testing was contrary to the basic philosophy of competency-based instruction. Other instructors disagreed and felt that testing was necessary in order to assure acquisition of required prerequisite concepts. Most students objected to testing for obvious reasons, while a few honest types admitted that they felt that this approach did result in greater learning. It was finally decided that the only reasonable way to settle the question was to conduct an experimental study in order to determine whether testing did or did not result in higher levels of student learning.

---

[2]Some enablers do not lend themselves to a testing situation. For example, the enablers for a task requiring the students to write a review of related literature involve the identification and annotation of at least 15 references related to a selected problem.

Prior to the start of the Fall Quarter, 1974, all students enrolled in the research course (N=126) were randomly assigned to one of three treatment groups: Group 1 - written exercises only; Group 2 - short answer tests; Group 3 - individual choice. A 1-way ANOVA indicated that there was a significant difference among the three groups on a 35-item "surprise" final examination. Application of Scheffé's multiple comparison procedure indicated that Group 2 retained significantly more concepts than both Group 1 and Group 3. In fact, the group which took tests retained 25% more than the group which submitted written exercises ($\bar{X}$ = 20.70 versus 16.50) and 44% more than the group which had a choice ($\bar{X}$ - 20.70 versus 14.33). It is to be noted that no students in the choice group chose to take tests; thus students in this group were treated essentially the same as students in the written exercises group. Attitudinal data indicated that students in the testing group felt that, although they had had to study more, they had learned more (see Appendix A for a more complete description of the study). Therefore, it was concluded that testing did result in higher levels of retention. Since the completion of the study, testing has been a required component of the course.

Grading

From the Fall Quarter of 1972 through the Winter Quarter of 1974, all grading was on a credit-no credit (CR-NC) basis. Within -course performances on tasks and enablers were each reported as being credit- level or not. The grade of CR reflected achievement of criteria regardless of the time required to reach criteria or the number of attempts involved. A course grade of CR indicated achievement of CR on all tasks and enablers within the course. Increasingly during this period, objections were raised by both instructors and students, on both philosophical and practical grounds.

One major philosophical issue was whether a student who demonstrates a competency on say a seventh attempt is equally competent as a student who demonstrates a competency on a first attempt; the majority opinion was a definite "No!" Another issue was whether level of performance, or competence, can be or should be assessed and reported; the majority opinion on this issue was "Yes!" on all counts. Practically speaking, the CR-NC reporting system was found to be unacceptable to many employer groups and graduate schools. Since every graduate's transcript looked virtually the same in terms of grades, selection of individuals for employment or graduate admission on the basis of achievement was basically impossible. Provision of supplementary data, such

as letters of recommendation, usually did not compensate for deficiencies in the grading system. As a result, our graduates were in many instances penalized.

In the Spring Quarter of 1974, the grade of honors credit (HC) was adopted. The addition of the HC grade was clearly a half-way measure and only a partial solution to the problem. FIU, however, was bound by an agreement with the Florida Board of Regents to try the CR-NC system for at least three years. Introduction of the HC grade was the only measure that could be taken until the Fall Quarter of 1975. In this quarter FIU officially adopted an A-B-C-D-NC grading system at both the undergraduate and graduate levels.[3] It was rather humorous to note the number of students who solicited clarification on the "new" grading system.

It is within the above-described historical perspective that the research course has operated. Although the grading system currently involves letter grades, evaluation is still criterion referenced. At the beginning of each quarter the criteria for each grade are presented to the students enrolled in the research course (see Appendix B). A grade of "B" indicates demonstration of competency at an acceptable level. A point system is utilized such that each task and set of

---

[3]Current indications are that within some period of time (as yet unspecified) the grade of NC will be replaced by the grade of F.

enablers is worth a certain number of points and final course grade is based on an accumulated total. While some tasks and enablers are graded as CR or NC, 4 major tasks are worth 20 points each and 7 sets of enablers (6 of which are evaluated through short-answer tests) are worth 10 points each. A student who minimally meets criteria on a task receives 16 points; 20 points are given only for essentially flawless performance. Other point values reflect varying degrees of acceptability. With respect to tests, 8 points is considered acceptable. The course is worth a total of 150 points and grades are assigned according to accumulated points. A grade of "B", for example, requires a minimum of 120 points which corresponds to an average of 16 points per task and 8 points per set of enablers (4 x 16= 64, 7 x 8=56, 64 + 56= 120).

Although it is basically graded CR-NC, the final task in the course gives students who lack a few points for a given grade one more opportunity to demonstrate competence. This task represents evaluation of one of the course's basic goals - to produce competent consumers of research - and requires students to read and evaluate a published research report. Evaluation of the report includes, for example, identification, description, and analysis of the experimental design used in the reported study. A student who lacks a few points for a "B" grade in the course, who does well on the final task, is given the "B" grade. Such a student

demonstrates that although performance on certain previous tasks may have been less than acceptable, competence is now evident. Thus, in essence, enablers are prerequisites for tasks and all preceding tasks are prerequisites for the final task, evaluation of a research report.

Pacing

Originally, progress through the course was completely student-paced although there were (and still are) regularly scheduled lectures. Students attempted tasks and enablers on dates and at times convenient to them. Each week, individual feedback could be received at any time during 20 regularly scheduled office hours including one evening. The intent was to meet student needs and to give students maximum flexibility. At the end of the quarter, however, only 34% (12 out of 35) students completed the course. The remaining students had an indefinite period of time in which to complete course requirements; a number of students did so during the following quarter, some never did. During the next five quarters, the end-of-quarter completion rates fluctuated between 38% and 64%. The burden on the instructors increased from quarter to quarter; each new quarter an instructor had responsibility for a group of new students and an increasing number of "carry-over" students from previous quarters. It was evident that many students could not deal effectively with the freedom and flexibility permitted them.

Beginning with the Fall Quarter of 1974, and during succeeding quarters, instructors increasingly imposed time restrictions on the achievement of tasks and enablers. More and more assignments were required to be completed by pre-specified dates; tests, for example,were administered in class on scheduled dates. By the Fall Quarter of 1975,(the quarter in which letter grades were introduced), all tasks and enablers were required to be completed by instructor-determined dates. The result was a progressively higher percentage of students completing the course by the end of a given quarter (see Figure 1). As Figure 1 indicates, completion rates progressively increased to 68%, 79%, 95%, and 100%. During the last 3 quarters, completion rates have fluctuated between 95% and 100%. Thus, elimination of self-pacing has benefited both students and instructors and the end-of-quarter completion rate has risen from 34% to 95-100%.

Recycling

Originally, unlimited recycling was permitted. A student could attempt any task or set of enablers as many times as necessary in order to reach criterion. Each unsuccessful attempt was followed by individual feedback provided by the instructor. As a number of students recycled as many as five or six times (the record being seven recycles for one task), the demands made on instructors were great. Since there was no penalty attached to a recycle, unprepared

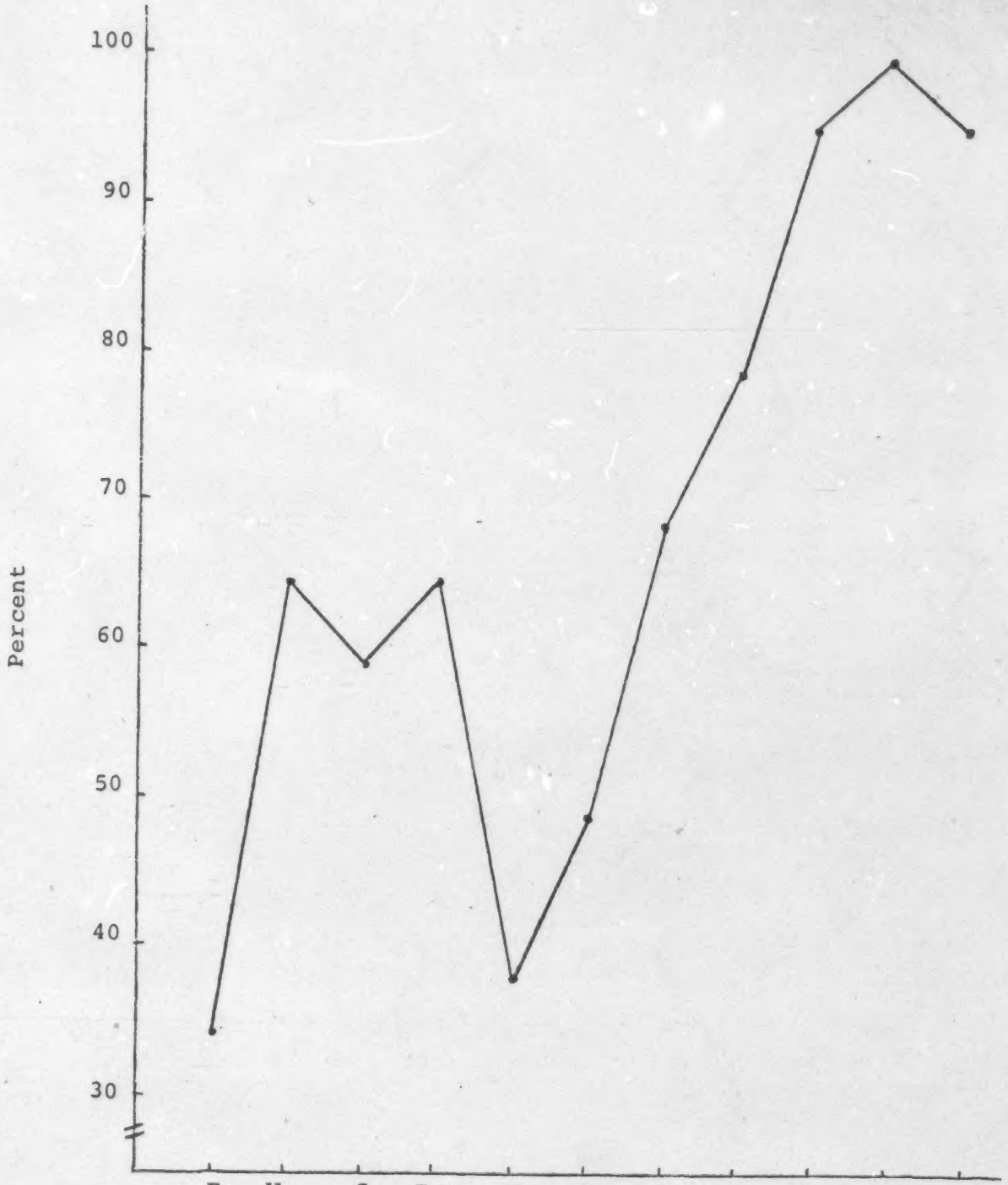

Figure 1. Percentage of students completing a competency-based research course by the end of each of 11 quarters (excluding Summers), Fall, 1972-Winter,1976.

Note: If appropriate, data points represent combined data for multiple sections.

students had nothing to lose by attempting a task. The "might as well take a shot at it" syndrome was even more of a problem when tests were introduced; some students would take a test, achieve a low score, receive feedback, and immediately retake the test. A number of students would devote entire afternoons to taking and retaking a test until they reached criterion. It become evident that limits had to be placed on recycling.

During the quarter in which restrictions began to be placed on self pacing, time restrictions were also placed on recycling. Basically, a student had to wait a minimum of two days (with a maximum of seven days allowed) before attempting a task or set of enablers previously attempted unsuccessfully. Not surprisingly, this simple procedure reduced the maximum number of recycles actually observed from seven to two.

With the introduction of letter grades, a policy was introduced such that excessive recycles result in a reduction of the maximum grade attainable. Currently, in order for a student to earn 17 or more points on a task (9 or 10 on a test), he or she must do so on the first attempt. Two further recycles are permitted (one for a test), but the maximum point value becomes 16; additional recycles are allowed but the maximum point value attainable becomes progressively lower (see Appendix B). As noted previously, despite the options, no student has recycled more than twice since grades were introduced.

A derivative benefit of the restrictions on recycling and the point losses associated with recycling has been an increase in the quality and productiveness of feedback sessions.

Students are considerably more intersted in determining their areas of deficiency and in correcting same as efficiently as possible; the "might as well take a shot at it" syndrome has virtually disappeared.

Conclusions

Based on nearly four years of experience, student performance data, and research findings, the following conclusions have been drawn concerning the components of a competency-based introductory research course:

1. Prespecified competencies and subcompetencies are essential and desirable components of a competency-based research course. If carefully formulated and reviewed, they are likely to remain relatively stable over time.

2. Criterion-referenced evaluation is also an essential and desirable component of a competency-based research course. In order for criterion-reference evaluation to be effective, students must be provided with precise, realistic models of expected performance.

3. Testing of subcompetencies is apparently essential in order to assure comprehension of prerequisite knowledge and skills. Testing has resulted in significantly greater (statistically and practically) retention of research concepts as compared to the mere submission of written exercises.

4. Credit - no credit grading is unacceptable, on both philosophical and practical grounds, to students, instructors, employees, and graduate schools. The results of criterion-referenced evaluation can be reported with a letter grading system. Grades of A,B,C,and D can be used to report demonstration of varying degrees of competence, with a B (at the graduate level) representing an acceptable level. A point system is an efficient, though not necessary, approach to such grading.

5. Self pacing is detrimental to both students and instructors. Data indicate that increased instructor pacing can increase course completion rate from a low of 34% to at least 95%.

6. Unlimited recycling is inefficient and unproductive, even when accompanied by intensive, individualized feedback. Placing restrictions on recycling and specifying consequences for excessive recycling result in more efficient achievement of competencies and increase the effectiveness of feedback.

APPENDIX A

Comparative Effectiveness of Tests Versus Written Exercises in a Competency-Based Research Course

LORRAINE R. GAY
PAUL D. GALLAGHER

Florida International University

Lorraine R. Gay
School of Education
Florida International University
Miami, Florida 33199

(a) Objective(s)

The purpose of this study was to determine whether required tests would result in greater retention of research concepts and procedures than required written exercises. It was hypothesized that students in an introductory educational research course who are required to pass a series of tests will exhibit significantly greater retention of the tested concepts and procedures on a final examination than students required to successfully complete a series of written exercises on the same concepts.

(b) Prospectives

The desirability of testing as a method of evaluation has been increasingly questioned. It has been suggested that the "cramming" phenomenon generally associated with testing results in meaningless memorization of information which is quickly forgotten. It has also been stated that the anxiety generated by testing may actually have a detrimental, interfering effect on the intended learning. As a result of these beliefs a trend toward utilization of alternate evaluation strategies has been developing. This trend is especially visible in the competency-based education approach with its emphasis on behavioral demonstration of performances and its deemphasis of written tests. While actual performance of competencies may indeed be the optional evaluation strategy, not all concepts and procedures can be so evaluated. There are a number of prerequisite concepts and procedures, acquisition of which facilitate achievement and demonstration of higher-order learnings (competencies), which do not lend themselves to performance evaluation. Even if they did, the large number of such prerequisites would make individual performance evaluation infeasible. Since these concepts and procedures are prerequisites and therefore necessary, however, their acquisition must be assured in an effective and practicable manner.

(c) Method

At the beginning of the fall quarter, 126 graduate students enrolled in an introductory educational research course were randomly assigned to one of three treatment groups and to one of

three instructors. Thus each treatment group initially comprised 42 students and each instructor was asisgned 42 students equally representing each of the 3 treatment groups. No pretest was administered since pretesting in 3 previous quarters had shown that students entering the course are completely lacking in research knowledge and skills. Students in all treatment groups received the same large-group lectures and all students were required to demonstrate the same competencies. Acquisition of prerequisite concepts and procedures, however, was demonstrated by different students in different ways, depending upon group membership:

Group 1 - Students demonstrated acquisition of all prerequisites through the submission of written exercises.

Group 2 - Students demonstrated acquisition of some prerequisites through submission of written exercises, but for most prerequisites through the passing of a series of short-answer tests.

Group 3 - Students had their choice; for any given set of prerequisites they could submit them as written exercises or they could pass a short-answer test.

Students in all groups had to achieve a criterior of 100% on both written exercises and tests. Students in all groups also recycled (resubmitted exercises, retook tests) as many times as necessary in order to reach criterion. Complete self-pacing was permitted for all students; exercises were submitted and tests taken on dates and at times individually determined by each student. At the end of the quarter, during the last class meeting, all students were administered a "surprise" 35-item short-answer final examination as a posttest measure of retention. The examination measured only those concepts and skills which had been measured on the previous tests of prerequisites. Students were also asked several attitudinal questions.

(d) Data Source

As a result of normal attrition and the self-pacing element, a number of students did not complete the course by the time of the last class meeting. Only posttest scores of students who did complete the course by this date were included in data analyses. Since scoring was somewhat subjective due to the short-answer nature of the posttest, 20% of the tests were scored twice; 80% of the tests were scored once and then 20% of the items were re-scored. The resulting scoring error rate was found to be 1.6%. As a further assessment of scorer reliability, another instructor re-scored all of the posttests; inter-scorer reliability was calculated to be $r = .94$.

(e) Results and Conclusions

Posttest means for groups 1, 2, and 3 were 16.50, 20.70, and 14.33, respectively; standard deviations were 6.05, 5.40, and 6.78, respectively. A one-way analysis of variance indicated that there was a significant difference among the 3 groups ($F = 4.44$, $df = 2/45$, $p < .05$). Application of Scheffé's multiple comparison procedure indicated that Group 2 - testing retained significantly more concepts and skills than Group 1 - written exercises ($F = 4.10$, $df = 2/45$, $p < .05$), and Group 3 - choice ($F = 7.79$, $df = 2/45$, $p < .05$). There was no significant difference between the retention of Groups 1 and 3 ($F = .25$, $df = 2/45$, $p > .05$). All students in Group 3 chose not to take tests and thus actually behaved the same as students in Group 1. It was concluded that tests were more effective in promoting retention than written exercises. Attitudinal data indicated that students in Group 2 felt that although they had had to work harder, they had learned more.

(f) Educational Importance of the Study

This study suggests that testing as a method of evaluation, no matter how distasteful it may be to educators and students, does promote greater retention of learned concepts and procedures. The current trend toward the elimination of formal, written tests may therefore be counterproductive with respect to one of education's major goals.

APPENDIX B

Grading Policy for EDU 507

With the exception of those tasks and enablers which are graded Credit-No Credit, each task will be worth a maximum of 20 points and each set of enablers (primarily evaluated as enabler tests) will be worth a maximum of 10 points. Thus there is a possible total of 150 points to be earned in the course (20 points times 4 tasks = 80 points, 10 points times 7 enabler evaluations = 70 points). At the end of the course, grades will be awarded based on point totals as follows:

A = 140-150 points
B = 120-139 points
C = 105-119 points
D = 90-104 points
NC= less than 104 points

A total of three attempts (with a minimum of two days and a maximum of one week between each attempt) are permitted in order to achieve a score of 16 on a given task. Further attempts are permitted, but the maximum score becomes 14. Scores of 16 and above may not be increased.

A total of two attempts (with a minimum of two days and a maximum of one week between each attempt) are permitted in order to achieve a score of 8 on an enabler test. Further attempts are permitted but the maximum score becomes 7. Scores of 8 and above may not be increased.

Point Values for Course Requirements

TASKS		ENABLERS	
1-B	20 points	2-2&2-3=	10 points
2	20 points	Test 4 =	10 points
6	20 points	Test 5 =	10 points
7	20 points	Test 6(1)=	10 points
	80 points	Test 6(2)=	10 points
		Test 7(1)=	10 points
		Test 7(2)=	10 points
			70 points

80+70=150 total points